



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

Address: COMMISSIONER FOR PATENTS

P.O. Box 1450

Alexandria, Virginia 22313-1450

www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/779,402	02/13/2004	Steven J. McCarthy	ID-504 (80226)	2799
89137	7590	09/01/2010		
Allen, Dyer, Doppelt, Milbrath & Gilchrist - RIM				
255 S. Orange Avenue				
Suite 1401				
Orlando, FL 32801				
EXAMINER				
NGUYEN, MINH CHAU				
ART UNIT		PAPER NUMBER		
2445				
NOTIFICATION DATE		DELIVERY MODE		
09/01/2010		ELECTRONIC		

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

creganoa@addmg.com

portfolioprossecution@rim.com

Office Action Summary

Application No.

10/779,402

Applicant(s)

MCCARTHY ET AL.

Examiner

MINH-CHAU NGUYEN

Art Unit

2445

Period for Reply -- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 15 April 2010.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-3,5,6,8-10,12,14,15,17,18,20 and 22-28 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-3,5,6,8-10,12,14,15,17,18,20 and 22-28 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 13 February 2004 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-940)
- 3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

DETAILED ACTION

A request for continued examination under 37 CFR 1.114, including the fee set forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this application is eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on 04/15/2010 has been entered.

This action is response to communications: application, filed on 02/13/2004; amendment filed on 04/15/2010. Claims 1-3, 5-6, 8-10, 12, 14-15, 17-18, 20 and 22-28 are presented for further examination.

The applicant's arguments filed on 04/15/2010 have fully considered and they are persuasive based on the applicant's amendment; however the arguments are moot in view of the new ground(s) of rejection. See rejection(s) below for details.

Claim Objections

Claim 17 is objected to because of the following informalities:

Claim 17 recites "A computer-readable medium having computer-executable instructions" should be "A computer-readable medium storing computer-executable instructions". Appropriate correction is required.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 17-18, 20 and 28 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 17-18, 20 and 28 recite "a computer readable medium". In light of the specification, there is no description of the computer readable medium. Thus, the claims as a whole covers both transitory and non-transitory medium. Therefore, the claim fails to fall within one of the four statutory categories of invention recited in 35 U.S.C. § 101: process, machine, manufacture, and composition of matter.

The claims may be amended by changing "a computer readable medium" to "a non-transitory computer readable medium", thus excluding that portion of the scope covering transitory media. Therefore, an amendment would limit the claims to an eligible embodiment.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-3, 5-6, 8-10, 12, 14-15, 17-18, 20 and 22-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Albert et al. (US 6,970,913) hereinafter "**Albert**", and in view of Dar et al. (US 2004/0205120) hereinafter "**Dar**".

Claim 1

Albert teaches a communications system comprising:

a plurality of servers [*i.e.* servers 221-223] connected together in a network [*i.e.* network 210] for processing a plurality of different job types [*i.e.* processing a plurality of task types for flowing of packets or handling connections (e.g. "routing the packet, gathering statistics...modifying packet" in col. 7, lines 14-18; and "Forwarding agents can accomplish their required tasks" in col. 8, line 11); and these task types will be implemented by a plurality of virtual machines on the servers] having respective resource usage [*i.e.* processing capacity usage] associated therewith (Albert, figure 2A; col. 6, line 51-col. 7, line 30; col. 8, lines 8-12, lines 57-67; col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51);

each server, after beginning execution of at least one job, determining a respective health metric thereof based upon at least one job being executed thereby [*i.e.* "The nature of the feedback messages from the real machines is that the messages somehow express the level of load on the real machine as a result of handling connections" and "the server determines the usage of processing capacity for each of the virtual machines that is being implemented"

in col. 30, lines 1-3, lines 21-23, thus the usage of processing capacity is considered as the health metric which determined from the result after the task being executed] and weighting the health metric of the at least one job [*i.e. weighting the usage of processing capacity to obtain "a weight" for the virtual machine*] (Albert, col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51); and

said servers mapping the weighted health metrics to a common scale [*i.e. expressing the weights to a common level (e.g. high or low level)*] (Albert, col. 30, line 1-col. 31, line 3; col. 32, lines 14-51); and

a dispatcher [*i.e. service manager 1 or 1140*] for collecting the commonly scaled weighted health metrics [*i.e. weights or levels of load*] from said servers [*i.e. the servers the real machines*] by polling said servers for the weighted health metrics [*i.e. retrieving is considered as polling the weights of the real machines*] and distributing jobs to said servers based thereon [*i.e. assigning connection tasks to the real machines based on the weights*] (Albert, figures 2A, 11A; col. 6, line 51-col. 7, line 30; col. 8, lines 8-12, lines 57-67; col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51).

Albert fails to teach different resource usage characteristics; and the health metric is determined based upon the respective resource usage characteristic after execution of the job, the resource usage characteristic

representing resources being consumed by the at least one job; and the weighted health metrics for different resource usage characteristics.

However, in an analogous art, **Dar** teaches different resource usage characteristics [*i.e. processor (CPU), memory, and input/output (I/O) metrics*] (Dar, paragraphs 0029); the health metric is determined based upon the respective resource usage characteristic after execution of the job [*i.e. the health of the server is considered as the health metric which aggregating the metrics. Moreover, Dar discloses "This monitoring may be periodic, e.g., every 10 seconds...or a synchronous monitoring of a different period would be acceptable", thus the monitoring could be run for obtaining the health and/or the metrics after a period of time of the program execution*], the resource usage characteristic representing resources being consumed by the at least one job [*i.e. the metrics are consumed by the program*] (Dar, paragraphs 0001, 0027, 0029-0032); and the weighted health metrics for different resource usage characteristics [*i.e. the total loads for the servers for different resource (e.g. CPU, memory, I/O) loads*] (Dar, paragraphs 0001, 0027, 0029-0032).

Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate the features of different resource usage characteristics; and the health metric is determined based upon the respective resource usage characteristic after execution of the job, the resource usage characteristic representing resources being consumed by the at least one job; and the weighted health metrics for different resource usage

characteristics, as disclosed by Dar, into the teachings of Albert. One would be motivated to provide "one or more of the following advantages...to improve resource utilization by the servers...Network server load balancing may be improved. Availability and/or scalability of network servers can be improved" [Dar, paragraph 0016].

Claim 2

Albert in combination with Dar teach the communications system of Claim 1 wherein the resource usage characteristics comprise at least one processing utilization characteristic and at least one input/output utilization characteristic [*i.e. processor (CPU) and input/output (I/O) metrics*] (Dar, paragraph 0029).

Claim 3

Albert in combination with Dar teach the communications system of Claim 1 further comprising a knowledge base [*i.e. a database*] for cooperating with said dispatcher [*i.e. the service manager*] for storing the weighted health metrics [*i.e. storing the weights*] (Albert, col. 31, lines 49-59).

Claim 5

Albert in combination with Dar teach the communications system of Claim 1 wherein said servers provide completed job results to said dispatcher, and wherein the weighted health metrics are provided to said dispatcher with the completed job results [*i.e. "the feedback messages from the real machines is that the messages somehow express the level of load on the real machine as a result of handling connections", and "a process executed on a server for determining a weight to be sent to the service manager in a feedback message... to determining the weights..., in a step 1206, the server determines the remaining processing capacity", thus the weights are sent to the service manger for the completed job/process results*] (Albert, Col. 30, L. 1-49).

Claim 6

Albert in combination with Dar teach the communications system of Claim 1 further comprising at least one load generator [*i.e. load balancer*] for generating the jobs for said servers and communicating the jobs to said dispatcher; and wherein said dispatcher [*i.e. the service manager*] further provides the completed job results to said at least one load generator [*i.e. "load balancer to be located between group of servers. It would be advantageous if a distributed architecture could be used for load balancing and if a server manager could be used to gather feedback from servers to make better load balancing decisions" in col. 3, lines 60-64*] (Albert, col. 3, line 59-col. 4, line 3; col. 8, lines 57-67; col. 11, lines 56-65; col. 28, lines 9-47; col. 30, lines 1-49; col. 32, lines 14-51).

Claim 8

Albert in combination with Dar teach the communications system of Claim 1 wherein the jobs relate to electronic mail (e-mail) processing [*i.e. the programs provide email service*] (Dar, paragraph 0001). One would be motivated to provide a wide array of services to clients via the network [Dar, paragraph 0001]

Claim 9

Albert teaches a load distributor for a plurality of servers [*i.e. servers 221-223*] connected together in a network [*i.e. network 210*] for processing a plurality of different job types [*i.e. processing a plurality of task types for flowing of packets or handling connections (e.g. "routing the packet, gathering statistics...modifying packet" in col. 7, lines 14-18; and "Forwarding agents can accomplish their required tasks" in col. 8, line 11); and these task types will be implemented by a plurality of virtual machines on the servers*] having respective resource usage [*i.e. processing capacity usage*] associated therewith (Albert, figure 2A; col. 6, line 51-col. 7, line 30; col. 8, lines 8-12, lines 57-67; col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51); each server, after beginning execution of at least one job, determining a respective health metric thereof based upon at least one job being executed thereby [*i.e. "The nature of the feedback messages from the real machines is that the messages somehow express the level of load on the real machine as a result of handling connections" and "the server*

determines the usage of processing capacity for each of the virtual machines that is being implemented" in col. 30, lines 1-3, lines 21-23, thus the usage of processing capacity is considered as the health metric which determined from the result after the task being executed] and weighting the health metric of the at least one job [*i.e. weighting the usage of processing capacity to obtain "a weight" for the virtual machine*] (Albert, col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51), the load distributor comprising:

a dispatcher [*i.e. service manager 1 or 1140*] for collecting the commonly scaled weighted health metrics [*i.e. weights or levels of load*] from said servers [*i.e. the servers or the real machines*] by polling said servers for the weighted health metrics [*i.e. retrieving is considered as polling the weights of the real machines*] and distributing jobs to said servers based thereon [*i.e. assigning connection tasks to the real machines based on the weights*] (Albert, figures 2A, 11A; col. 6, line 51-col. 7, line 30; col. 8, lines 8-12, lines 57-67; col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51); and

said servers mapping the weighted health metrics to a common scale [*i.e. expressing the weights to a common level (e.g. high or low level)*] (Albert, col. 30, line 1-col. 31, line 3; col. 32, lines 14-51); and

a knowledge base [*i.e. a database*] for cooperating with said dispatcher [*i.e. the service manager*] for storing the weighted health metrics [*i.e. storing the weights*] (Albert, col. 31, lines 49-59).

Albert fails to teach different resource usage characteristics; and the health metric is determined based upon the respective resource usage characteristic after execution of the job, the resource usage characteristic representing resources being consumed by the at least one job; and the weighted health metrics for different resource usage characteristics.

However, in an analogous art, **Dar** teaches different resource usage characteristics [*i.e. processor (CPU), memory, and input/output (I/O) metrics*] (Dar, paragraphs 0029); the health metric is determined based upon the respective resource usage characteristic after execution of the job [*i.e. the health of the server is considered as the health metric which aggregating the metrics. Moreover, Dar discloses "This monitoring may be periodic, e.g., every 10 seconds...or a synchronous monitoring of a different period would be acceptable", thus the monitoring could be run for obtaining the health and/or the metrics after a period of time of the program execution*], the resource usage characteristic representing resources being consumed by the at least one job [*i.e. the metrics are consumed by the program*] (Dar, paragraphs 0001, 0027, 0029-0032); and the weighted health metrics for different resource usage characteristics [*i.e. the total loads for the servers for different resource (e.g. CPU, memory, I/O) loads*] (Dar, paragraphs 0001, 0027, 0029-0032).

Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate the features of different resource usage characteristics; and the health metric is determined based upon

the respective resource usage characteristic after execution of the job, the resource usage characteristic representing resources being consumed by the at least one job; and the weighted health metrics for different resource usage characteristics, as disclosed by Dar, into the teachings of Albert. One would be motivated to provide "one or more of the following advantages...to improve resource utilization by the servers...Network server load balancing may be improved. Availability and/or scalability of network servers can be improved" [Dar, paragraph 0016].

Claim 14

Albert teaches a job distribution method for a plurality of servers a plurality of servers [*i.e. servers 221-223*] connected together in a network [*i.e. network 210*], the servers for processing a plurality of different job types [*i.e. processing a plurality of task types for flowing of packets or handling connections (e.g. "routing the packet, gathering statistics...modifying packet" in col. 7, lines 14-18; and "Forwarding agents can accomplish their required tasks" in col. 8, line 11); and these task types will be implemented by a plurality of virtual machines on the servers*] having respective resource usage [*i.e. processing capacity usage*] associated therewith (Albert, figure 2A; col. 6, line 51-col. 7, line 30; col. 8, lines 8-12, lines 57-67; col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51), the method comprising:

determining a respective health metric of each server after it begins execution of at least one job based upon the at least one job being executed thereby [i.e. *"The nature of the feedback messages from the real machines is that the messages somehow express the level of load on the real machine as a result of handling connections" and "the server determines the usage of processing capacity for each of the virtual machines that is being implemented"* in col. 30, lines 1-3, lines 21-23, thus the level of load of each real machine is determined based on the usage of processing capacity from the result after the task being executed] and weighting the health metric of the at least one job [i.e. *weighting the usage of processing capacity to obtain "a weight"*] (Albert, col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51);

polling said servers for the weighted health metrics [i.e. *retrieving is considered as polling the weights of the real machines*] (Albert, col. 32, lines 14-51) and mapping the weighted health metrics to a common scale [i.e. *expressing the weights to a common level (e.g. high or low level)*] (Albert, col. 30, line 1-col. 31, line 3; col. 32, lines 14-51); and

distributing jobs to said servers based upon the commonly scaled weighted health metrics [i.e. *assigning connection tasks to the real machines based on the weights*] (Albert, col. 28, lines 9-47; col. 29, line 44-col. 30, line 49; col. 32, lines 14-51).

Albert fails to teach different resource usage characteristics; and the health metric is determined based upon the respective resource usage characteristic after execution of the job, the resource usage characteristic representing resources being consumed by the at least one job; and the weighted health metrics for different resource usage characteristics.

However, in an analogous art, **Dar** teaches different resource usage characteristics [*i.e. processor (CPU), memory, and input/output (I/O) metrics*] (Dar, paragraphs 0029); the health metric is determined based upon the respective resource usage characteristic after execution of the job [*i.e. the health of the server is considered as the health metric which aggregating the metrics. Moreover, Dar discloses "This monitoring may be periodic, e.g., every 10 seconds...or a synchronous monitoring of a different period would be acceptable", thus the monitoring could be run for obtaining the health and/or the metrics after a period of time of the program execution*], the resource usage characteristic representing resources being consumed by the at least one job [*i.e. the metrics are consumed by the program*] (Dar, paragraphs 0001, 0027, 0029-0032); and the weighted health metrics for different resource usage characteristics [*i.e. the total loads for the servers for different resource (e.g. CPU, memory, I/O) loads*] (Dar, paragraphs 0001, 0027, 0029-0032).

Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to incorporate the features of different resource usage characteristics; and the health metric is determined based upon

the respective resource usage characteristic after execution of the job, the resource usage characteristic representing resources being consumed by the at least one job; and the weighted health metrics for different resource usage characteristics, as disclosed by Dar, into the teachings of Albert. One would be motivated to provide "one or more of the following advantages...to improve resource utilization by the servers...Network server load balancing may be improved. Availability and/or scalability of network servers can be improved" [Dar, paragraph 0016].

Claim 22

Albert in combination with Dar teach the communications system of Claim 1, wherein the at least one job comprises e-mail delivery [*i.e. the programs provide email service, and the email delivery is included in the email service*] (Dar, paragraph 0001). One would be motivated to provide a wide array of services to clients via the network [Dar, paragraph 0001].

Claim 23

Albert in combination with Dar teach the communications system of Claim 6, wherein said at least one load generator [*i.e. load balancing*] comprises an e-mail aggregation engine [*i.e. the programs provide email service, and the load balancing*]

between the servers for performing the programs based on the aggregating metrics, thus the examiner interprets that the aggregating metrics would be done by an aggregation engine] (Dar, paragraphs 0001, 0016, 0029, 0033).

Claim 24

Albert in combination with Dar teach the communications system of Claim 6, wherein said servers also provide completed job results to said at least one load generator [i.e. *"load balancer to be located between group of servers. It would be advantageous if a distributed architecture could be used for load balancing and if a server manager could be used to gather feedback from servers to make better load balancing decisions"* in col. 3, lines 60-64] (Albert, col. 3, line 59-col. 4, line 3; col. 8, lines 57-67; col. 11, lines 56-65; col. 28, lines 9-47; col. 30, lines 1-49; col. 32, lines 14-51).

Claims 10, 12, 25, 26 are corresponding apparatus claims of system claims 2, 5, 22, 6. Therefore, they are rejected under the same rationale.

Claims 15, 27 are corresponding method claims of system claims 2, 22. Therefore, they are rejected under the same rationale.

Claims 17-18, 20, 28 are corresponding computer-readable medium claims of apparatus claims 9-10, 12, 25. Therefore, they are rejected under the same rationale.

Conclusion

The prior arts made of record and not relied upon are considered pertinent to applicant's disclosure:

1. Chellis et al., U.S. Patent No. 6,901,446 disclosed system and method for describing and automatically managing resources.
2. Colby et al., U.S. Patent No. 6,862,624 disclosed method and apparatus for directing a flow of packets based on request and server attributes.
3. Jones, U.S. PG-Pub. 2004/0267548 disclosed workload profiling in computers.

Correspondence Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to MINH-CHAU NGUYEN whose telephone number is (571)272-4242. The examiner can normally be reached on 7AM-3:30PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, VIVEK SRIVASTAVA can be reached on (571) 272-7304. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/MINH-CHAU NGUYEN/
Examiner, Art Unit 2445